UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF NORTH CAROLINA
WESTERN DIVISION
5:10-CT-3123-BO

SHAUN A. HAYDEN,		)
		)
	Plaintiff,	)
		)
		)
PAUL G. BUTLER,		)
		)
	Defendant.	)

EXPERT REPORT OF BRYAN GILBERT DAVIS

Punishment and Reprieve

Who gets parole in North Carolina?

by Bryan Gilbert Davis[1]

Contents

1. Introduction:

According to the North Carolina Department of Public Safety (DPS)[2], parole decisions should take into account factors including the nature and circumstances of the crime, the previous criminal record, prison conduct, prison program participation, as well as input from court officials, victims, and other

[1] Bryan Gilbert Davis's resume, outlining his experience and expertise, is attached.

[2] https://www.ncdps.gov/Index2.cfm?a=000003,002210,002212

interested parties. According to North Carolina state law, parole decisions can also relate to factors such as whether "His or her release at that time would unduly depreciate the seriousness of his or her crime or promote disrespect for law."[3] Thus, in looking at the letter of the law, one would expect parole release to correlate with a number of factors, such as the severity and nature of a prisoner's offense, his or her behavior while in custody, and his or her participation in work and enrichment activities, and the potential threat posed by a prisoner if released. Furthermore, one would hope that factors such as race, gender, or bias toward a particular parole analyst gender would play a minimal role in these decisions.

Fortunately, it is possible to test this hypothesis, and also examine other factors that may contribute to parole decisions in North Carolina. This is possible thanks to a public data set provided by the North Carolina Department of Public Safety, available on their website[4], which consists of several tables containing information on the prison population of North Carolina going back to 1972. This data includes fields pertaining to many of the factors mentioned above, including a record of prisoners' most serious offenses, their disciplinary infractions during their time in prison, escape attempts and prior convictions, their age, and their parole status. In addition, it also contains information pertaining to the gender, race, length of time incarcerated, and county of residence or incarceration. There are also fields indicating which parole analyst reviewed each case in preparation for the parole review.

Through statistical analysis of this dataset and consultation with lawyers from North Carolina Prisoner Legal Services, the author was able to conduct exploratory analysis on this data to understand the overall differences between paroled and non-paroled prisoners, and also construct a model to analyze which factors were the most influential in determining parole review outcomes. In order to focus the research question on a small number of cases, and eliminate possible interference from legal phenomena such as prison population caps, this analysis was aimed at the over 2,400 inmates given life sentences before 1995. As expected, the severity of an inmate's crime, behavior while in prison, and age were all found to be statistically significant factors in determining the likelihood of parole. However, analysis has also revealed that an inmate's sex and race are also influential factors, in and of themselves, in determining parole likelihood. In the sections below, this report will outline the relevant historical background, describe the methods and tools used to address the research question, and discuss the specific findings and their interpretation.

2. Historical Background:

The parole system in North Carolina has undergone numerous changes since its original inception in 1868. In its earliest form, the governor was empowered with the ability to make decisions regarding reprieves, commutations, and pardons, and this was expanded to include a system of supervised release. The governor or his staff retained this authority until 1955, when North Carolina established the state's earliest Parole Commission, which had exclusive authority to grant, revoke, and terminate parole. For the next 26 years, the Parole Commission had a great deal of discretion in making parole decisions, which sought to emphasize rehabilitation and public safety. However, in the 1980s, concerns about sentence disparities and a growing prison population gave rise to a new set of rules and standards. In 1987, the General Assembly passed the Prison Population Stabilization Act, known as the prison cap, which mandated that the Commission keep the prison population below a legally-determined level. This

[3] http://pardonandparole.uslegal.com/state-pardon-and-parole-laws/north-carolina/
[4] http://webapps6.doc.state.nc.us/opi/downloads.do?method=view

dramatically changed the parole process in North Carolina for the duration of its tenure, which ended in 1996. During this time, many inmates found guilty of misdemeanors were released categorically, without much consideration to their degree of rehabilitation or to public safety, as a way to prevent prison overcrowding. In 1994, the system changed yet again with the passage of the Structured Sentencing Act, which eliminated the parole system as it had previously existed, and removed the Commission's discretionary role for most crimes committed after October 1, 1994, with the exception being those incarcerated for driving under the influence.

This report aims to analyze the factors that influence the probability of being granted parole by the Commission for a certain class of offenders, namely those with life sentences convicted before 1995. By focusing on this select group of inmates, it is possible to limit the influence of the changing legal environment. First, by choosing only those prisoners who were convicted prior to 1995, we can be sure that the prison population we are analyzing was and is subject to the Parole Commission's discretion. Second, by focusing our analysis on those prisoners with life sentences, invariably guilty of serious felonies, we can be confident that such prisoners would not have been subject to any categorical release programs as a way to address prison overcrowding.

3. Methods:

This goal of this research is to analyze the factors that affect parole decisions for serious offenders in the North Carolina prison system. As mentioned in the historical summary, only those inmates whose crimes were committed before October 1, 1994 continue to be subject to parole review. Prisoners whose sentencing occurred before this date and who are still in prison, nearly 20 years, have invariably been found guilty of egregious crimes, predominately rape and murder. However, other prisoners also guilty of such crimes have been paroled and released. The goal of this research is to analyze the distinguishing factors between these two groups, and determine which factors have the most significant influence.

Generally, the goal of statistical analysis is to use patterns in data to determine causal relationships between different elements in the data. One crucial step in this process is variable selection, the process by which the researcher selects which elements to include in the dataset and which to exclude. For this particular research, the goal is to determine which factors affect the likelihood of parole. Parole decisions are discretionary, and thus subject to an incredible array of variables having to do with the parole commission itself. For instance, research has shown that parole decisions by judges are highly correlated with the time of day, before or after a break, that the decisions are made.[5] Another potentially influential factor is the political environment under which parole decisions are made: is there a political mandate for the governor to be "tough on crime", for instance? This report does not address these factors, and thus treats each parole decision as an independent decision made under similar conditions.

With this in mind, the researcher had to shape his questions to the data available. The data available on the DPS website contained a number of different tables, including:

[5] http://www.pnas.org/content/108/17/6889

Probation and Parole Client Profile, Impact Scheduling Request, Inmate Profile, Sentence Computations, Parole Analyst Review, Disciplinary Infractions, Financial Obligation, Offender Profile, Court Commitment, Sentence Component, Special Conditions and Sanctions, and Warrant Issued[6]

Each table was saved as a fixed-field-length .dat file, with the length of each field for each table listed in files containing field definitions that accompanied each .dat file. These accompanying field definitions allowed for the researcher to understand the overall design of the database and how each table fit together. Each prisoner is distinguished through a unique identification number, which is used to establish the link between each prisoner and his/her disciplinary record, sentence details, and so on. The data thus appeared to have many standard features of a relational database and could be queried as such if reconstructed correctly. This was accomplished, allowing the data to be accessed, searched, and aggregated so as to accomplish the research goals.

3.1. Variable Selection

Based on consultation Elizabeth Simpson and Mary Pollard at North Carolina Prisoner Legal Services, as well as analysis of North Carolina law regarding the parole decision-making process, it was clear that parole decisions should be based on factors including criminal record, disciplinary record while in prison, and age/health. The last field is relevant due to its impact on the threat of a certain individual to society, i.e. it is harder for somebody old or frail to commit violent crimes. In addition, given the general failure of social and governmental systems to eliminate racial and sexual bias, it was clear that fields pertaining to demographic information also needed to be included. Finally, it was crucial to have a field indicating whether or not an inmate had been paroled.

To this end, the researcher determined that the most relevant fields for this analysis were to be found in the following tables: Inmate Profile, Disciplinary Infractions, and Sentence Component. These tables had enough information to analyze several different factors that would potentially have an effect on parole decisions. These factors, explored and explained in the following section, were: age, sex, and race, length of incarceration, most serious crime, behavior in prison, escape attempt record, prior conviction record, and parole case analyst. This set of data covered almost all of the information that we felt should be included, with the exception of any record of the health status, which would serve as a potentially influential factor in parole decisions. However, to a limited degree, the age of an inmate can be a proxy for health, and thus health status is represented to some degree.

Desired frame of analysis	Corresponding table	Corresponding field
Demographic indicators	Inmate Profile	Age, race, sex
Criminal Record	Inmate Profile	Most serious conviction, prior flag
Behavioral/disciplinary record	Inmate Profile, Disciplinary	Infraction record and infraction

[6] http://webapps6.doc.state.nc.us/opi/downloads.do?method=view

	Infractions	coefficient, escape flag, length of incarceration
Parole commission bias	Inmate Profile	Parole analyst
Parole status	Inmate Profile	Administrative status, type of last update
Health indicators[7]	Inmate Profile	Age

3.1.1. Demographic Features

The most self-explanatory of fields are the demographic features. The analysis included inmates' race, gender, and age as possible factors determining parole. Out of these three, age is the least obvious. For age to be a factor in parole decisions, the field should reflect the age of the inmate at the time of the decision, not at the present day. So, for instance, if an inmate was paroled ten years ago, the age should reflect his/her age at the time of that decision. Fortunately, there is a field indicating the "date of last movement", which is the last date at which the prisoner's status changed, either via inter-prison transfer, or being granted parole, or terminating parole. There is still some room for error in this calculation however, as if the last date is the date at which a prisoner terminated parole, it is not the same date at which the prisoner was granted parole. There may be a difference of several years between these dates. Nonetheless, the time between a prisoner's birthdate and their "date of last movement" are a suitable approximation of age at the time of parole.

3.1.2. Criminal Record

The dataset contains information pertaining to the most serious crime that an inmate has been convicted of, and a yes/no indicating whether or not the inmate has any prior convictions. This research did not include a record of how serious those prior convictions were, or how many there were. Also, because the Inmate Profile table only lists the most serious offense for the current incarceration period, the researcher was unable to determine whether an inmate was serving a sentence for multiple convictions, for instance as a result of a robbery turned murder, where the inmate would have been charged with both.

3.1.3. Behavioral/Disciplinary Record: Escape Flag and the Infraction Coefficient

The Disciplinary Infractions contains a thorough record of all of the disciplinary infractions committed by inmates in the course of their incarceration, going back to the 1950s. Each record in the table is linked to a specific inmate via their ID number, and other fields contain information such as the type infraction, the date on which it occurred, whether or not the inmate was found guilty of the infraction, and so on. Obviously, some inmates have multiple infractions, and some inmates have very few or even none at all. As a way to measure the

[7] The DPS internal database does contain other health records, but these are not available in the public-facing tables. Such information can ostensibly be accessed by researchers if they submit an application requesting access, but two separate submissions of this application received no response from DPS.

behavior of inmates, the researcher developed a numerical measure of each prisoner's disciplinary record. This was done by first coding each type of infraction (77 in total) by order of severity on a scale from 1 to 3, with 1 being the least serious and 3 being the most serious. For example, the infraction of "verbal threat" was given a severity rating of 1, while "escape" or "threaten to harm-injure staff" were both given a severity rating of 3. Then, based on these severity scores, each prisoner's infraction coefficient was calculated as follows:

$$infraction\ coefficient = \frac{1 * (no.\ level\ 1\ ) + 2 * (no.\ level\ 2\ ) + 3 * (no.\ level\ 3\ )}{no.\ of\ years\ incarcerated}$$

The dividing by number of years allows the coefficient to summarize the behavior of the inmate over the course of their incarceration. Thus a prisoner with one verbal threat and one escape would have a numerator of 4, and if he/she had been in prison for 2 years would have a total infraction coefficient of 2. The weakness of this approach is that it assumes that the difference in severity between two infractions is constant and thus may underemphasize particularly egregious infractions. Adjusting the breadth of the severity scale, for instance by increasing it to a ten point scale, could ameliorate this issue. The current scale is provided in the appendix.

In addition to this infraction coefficient, the tables also contain a yes/no indicator of whether or not an inmate has attempted escape, which can serve as a unique identifier emphasizing the seriousness of escape attempts in determining parole decisions.

3.1.4. Parole Analyst Bias

One other field that the researcher wished to address was whether or not parole decisions were based upon bias toward a specific or select group of parole analysts. One field in the Inmate Profile table proved to be an ideal way to measure this influence, as each inmate was assigned to one of the 33 parole analysts, each indicated by an alphanumeric code. It is not clear whether these codes represent a specific individual parole analyst, a particular locale's analysts, or a select group of analysts. Nonetheless, it does not detract from the analysis to include this factor.

3.1.5. Parole Status

The goal of this analysis is to examine the factors that influence parole decisions, so a clear indicator of parole status was absolutely crucial. Initially, communication with representatives from DPS seemed to indicate that the only clear indicators of parole status were held in non-public tables only accessible to researchers through the submission of a research application. However, the Inmate Profiles table proved to contain two fields indicating parole status (neither was labeled "parole status", so two fields were used to confirm against the other). These were, respectively, a field indicating the "type of last inmate movement", which indicated whether the last movement had been parole, termination of parole, or otherwise, and "inmate admin. status code", which indicated whether an inmate was still active in the system, for which paroled inmates were marked "inactive".

3.1.6. Health Indicators

Of the factors that an analysis of parole decisions should attempt to account for, almost all are covered in some capacity through those elements outlined above. One that it is not covered, however, is the health of inmates. It is not exactly clear what influence the health of an inmate would have on parole decisions, but it is reasonable to assume that declining health substantially limits the risk posed by an inmate upon release. However, all information pertaining to the health status of inmates is stored in non-public tables in the DPS database, and thus cannot be accessed without applying via a research application. The author submitted this application on two separate occasions, but did not receive any feedback from DPS. As a result, this analysis does not include any direct information pertaining to the health of inmates, other than that which can be assumed to be generally associated with age. As prisoners age, they are naturally more likely to encounter health issues, and thus age can serve as a very rough proxy of health-related issues. Unfortunately, this must suffice for the time being.

3.2. Weaknesses of this Methodology

In terms of design, in addition to the inability to access relevant health information and the other small issues explained in the section above, the largest weakness in this research is the inability to analyze the parole process on a decision-by-decision basis. The current design only looks at the most recent decision regarding inmates' parole status, and thus ignores the history leading up to the most recent update. As a result, the analysis fails to account of several factors that a more ideal model would possess. For instance, the current style of analysis fails to account for changes in inmates' behavior over time. It also fails to record or account for the parole decisions leading up to the most recent decision, or for cases in which prisoners were given parole but re-arrested for new offenses. A more rigorous model would look at a similar set of variables as those mentioned above, but it would examine the status of these indicators *at the time of each parole review*, as opposed to for each person. Thus a single inmate might have dozens of instances which could provide data, as opposed to only one. The author does not feel that these weaknesses invalidate the strength of the research in this report, or the impact of the findings, but does acknowledge that there is room for further refinement of this process.

4. Analysis and Findings

The core of this report is an attempt to isolate the main differences between inmates who are granted parole and those that are not. This analysis will start with a descriptive account of the inmate population, emphasizing the overall differing characteristics of those who are granted parole versus those that are not. Second, this report will introduce the statistical model, utilizing a statistical technique known as logistic regression, which may elucidate possible causal relationships between different factors.

4.1. Descriptive Analysis

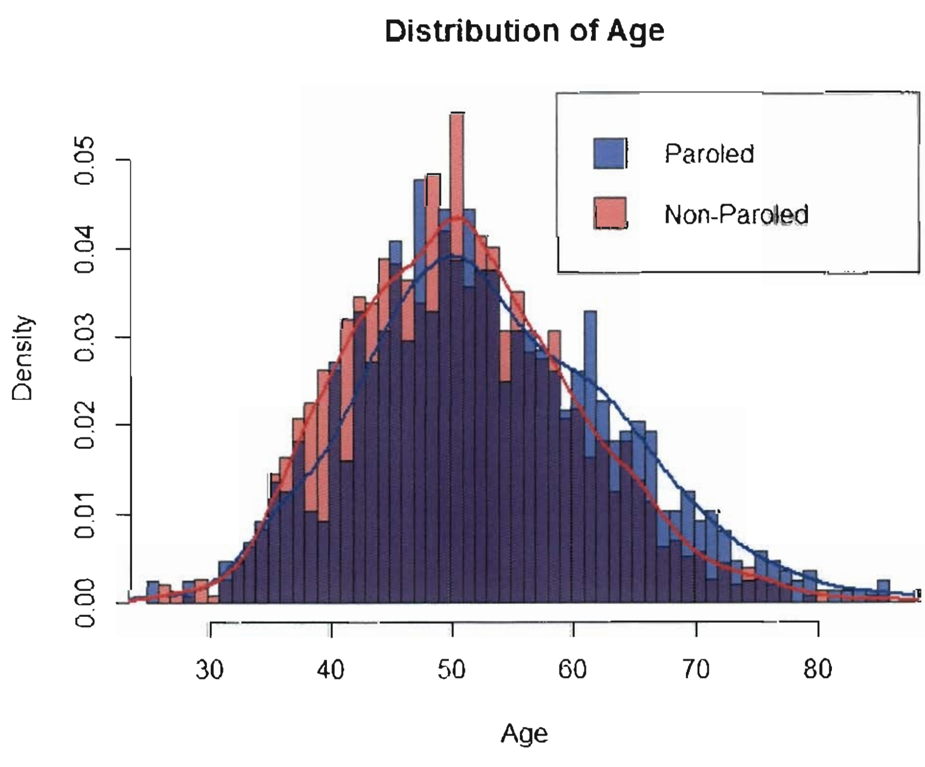
As mentioned above, this report considers only those inmates with life sentences who have been listed as "deceased". Altogether, this limitation restricts our analysis to the 2,466 inmates that meet these criteria. Within this group, paroled versus non-paroled offenders are distinguished by their "administrative status" and "type of last movement" fields. In the final count, 877 of these were paroled

and 1,589 were not, producing a parole percentage of 35.6%. The defining features of these two groups will be explored below, with the aid of tables and graphs.

4.1.1 Demographics

First and foremost, it makes sense to examine the demographic differences between those who have been granted parole and those who haven't. As can be seen in the graph below, parolees are, on average, older than non-paroled inmates, with an average and median age of 59 and 58 years old, respectively. Meanwhile, non-paroled offenders have a median and average of 53 years old. Thus it does appear that inmates who are older are, on average, are more likely to be granted parole. It is also likely that this difference could be partially accounted for by inaccuracies in the release times due to the lack of distinction between parole dates and parole termination dates.

In the graph below, histograms of the ages of paroled and non-paroled offenders are overlayed to emphasize the difference between the two. The purple space indicates where the two populations overlap. Furthermore, the red and blue lines are a kernel density approximation, illustrating the estimated prevalence of various ages as a continuous line as opposed to individual columns. The curves show clearly the trend of paroled offenders being comparatively older on average.



The racial breakdown of the prison population is fairly bipolar, dominated by white and black offenders, with a small number of Asian, Indian (Native American), and Other/Unknown. In

terms of sex, the population is overwhelmingly male, with only 3% female offenders. The table below allows comparison of the parole rates for specific demographic groups. It can be seen that, while black males represent 57% of the overall population, they represent only 53% of the paroled population. Meanwhile, white offenders are paroled at a rate almost exactly representative of their proportion in the overall population (37-38%). Thus there is a 4-5% discrepancy in the parole rates of black and white offender. In addition, women, only 3% of the population, represent 5% of the paroled population. It is quite possible that differences in both of these categories could result from other factors, including crime, disciplinary record, and so on. Thus these differences are not enough to prove that there is any systematic discrimination, but they do serve as foundations from which to ask how important race and gender are in determining parole outcomes.

Overall, it is clear that the parole rates if different groups differ substantially, although the low number of Asian, Indian, and Other/Unknown inmates makes analysis for these groups less fruitful.

Prison Population Demographics

1. TOTAL POPULATION

	Asian	Black	Indian	Other	Unknown	White	SUM
Female	0	29	1	0	0	46	76
Male	1	1409	61	30	1	888	2390
SUM	1	1438	62	30	1	934	
	Asian	**Black**	**Indian**	**Other**	**Unknown**	**White**	**TOTAL**
Female	0%	1%	0%	0%	0%	2%	3%
Male	0%	57%	2%	1%	0%	36%	97%
TOTAL	0%	58%	3%	1%	0%	38%	

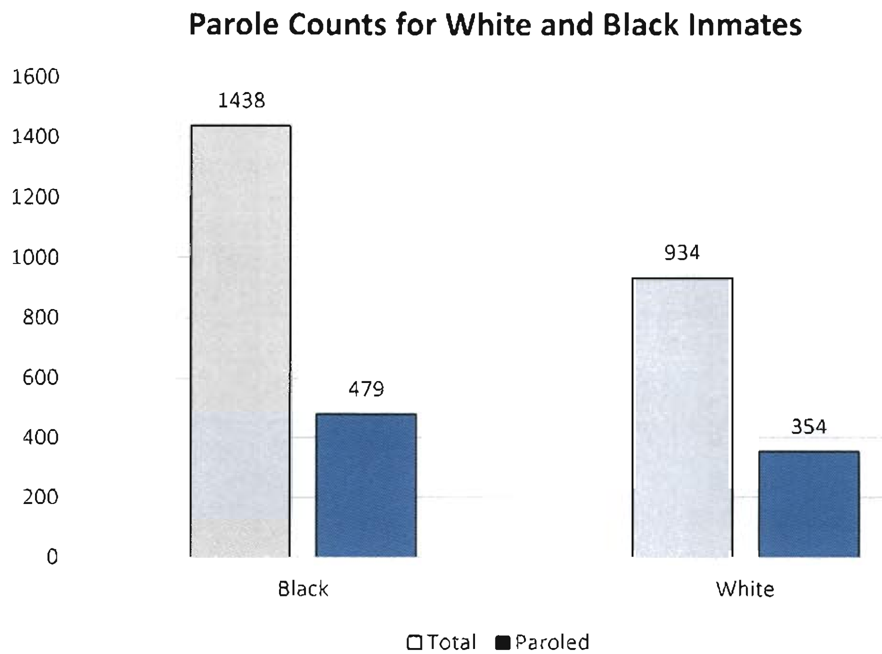
2. PAROLEE POPULATION

	Asian	Black	Indian	Other	Unknown	White	SUM
Female	0	17	1	0	0	27	45
Male	1	462	26	16	0	327	832
SUM	1	479	27	16	0	354	
	Asian	**Black**	**Indian**	**Other**	**Unknown**	**White**	**TOTAL**
Female	0%	2%	0%	0%	0%	3%	5%
Male	0%	53%	3%	2%	0%	37%	95%
TOTAL	0%	55%	3%	2%	0%	40%	

3. PERCENTAGE PAROLED

	Asian	Black	Indian	Other	Unknown	White	TOTAL
Female	NA	59%	100%	NA	NA	59%	59%
Male	100%	33%	43%	53%	0%	37%	35%
TOTAL	100%	33%	44%	53%	0%	38%	

With the vast majority of the prison population made up of white and black offenders, it is reasonable to exclude the other groups from the analysis. Below is a more direct comparison of the parole rates of white and black prisoners:



4.1.2 Behavioral/Disciplinary Record

One obviously important factor to be considered in parole decisions is the behavior of offenders while in prison. Common sense would have it that prisoners who commit many offenses would be less likely to be granted parole. The data does indeed bear this out.

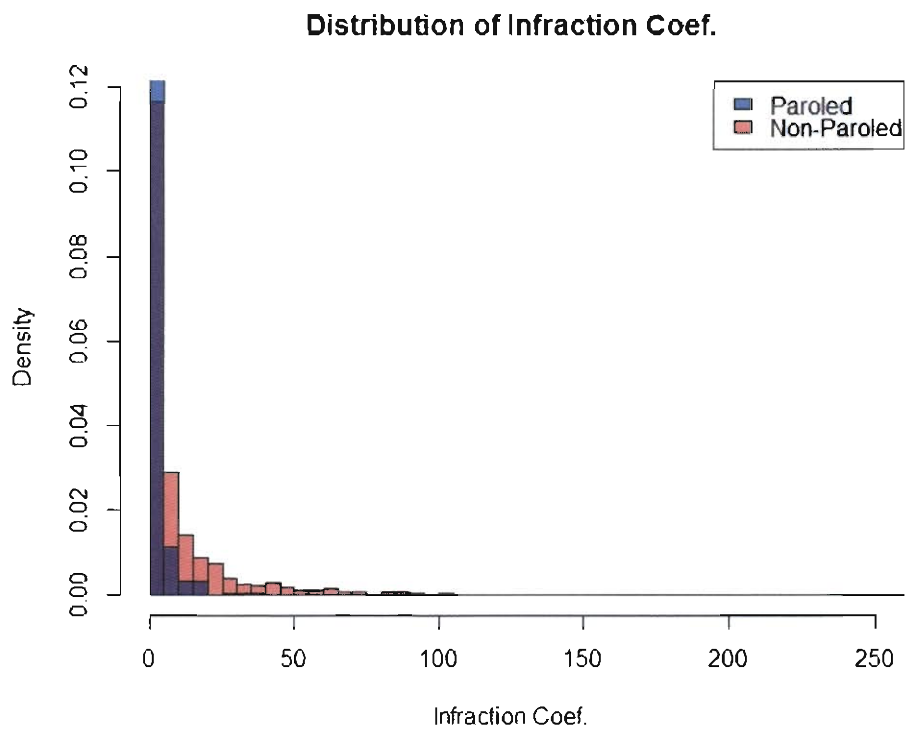
The chief measure of this disciplinary record in prison used in this report is the "infraction coefficient" outline in Section 3.1.3. This measure is based on the number and severity of disciplinary infractions, with each infraction graded on a 1 to 3 scale (from least serious to most serious), and the total infraction coefficient resulting from the sum of all these severity scores divided by the total number of years incarcerated.

To be clear, both there are both paroled offenders with relatively high infraction coefficients and non-paroled offenders with very low infraction coefficients, but there is nonetheless a clear pattern of paroled offenders having lower infraction coefficients. This is most visible in the differences in the mean and median. One might notice that for each group, the mean is significantly higher than the median. This is a result of the heavily skewed nature of the data. There are a few offenders with extremely high infraction coefficients, but the vast majority of offenders have very low infraction coefficients.

Infraction Coefficient						
	Minimum	1st	Median	Mean	3rd	Max

		Quartile			Quartile	
Total Population	0	0	1.5	8.8	7.1	825.9
Paroled	0	0	0.4	2.0	1.8	79.9
Non-paroled	0	0.1	3.1	12.4	11.4	825.9

The relationship can be explored with more clarity through the histogram below. Although both the paroled and non-paroled offenders are concentrated at the low end of the spectrum, is it clear that non-paroled offenders are less so. Very few paroled offenders have an infraction coefficient above 20.



4.1.3 Crime

Those offenders serving life sentences tend to be guilty of very serious offenses. The table below breaks down the inmate population by crime and parole status, sorted by the number of inmates in each crime category. Murderers and rapist make up the vast majority of the offender population, but there are also those guilty of property and drug crimes. Upon first inspection, it is immediately clear that the parole rates for different crime types vary widely. Sexual offenders appear to be paroled at a much lower rate than violent offenders.

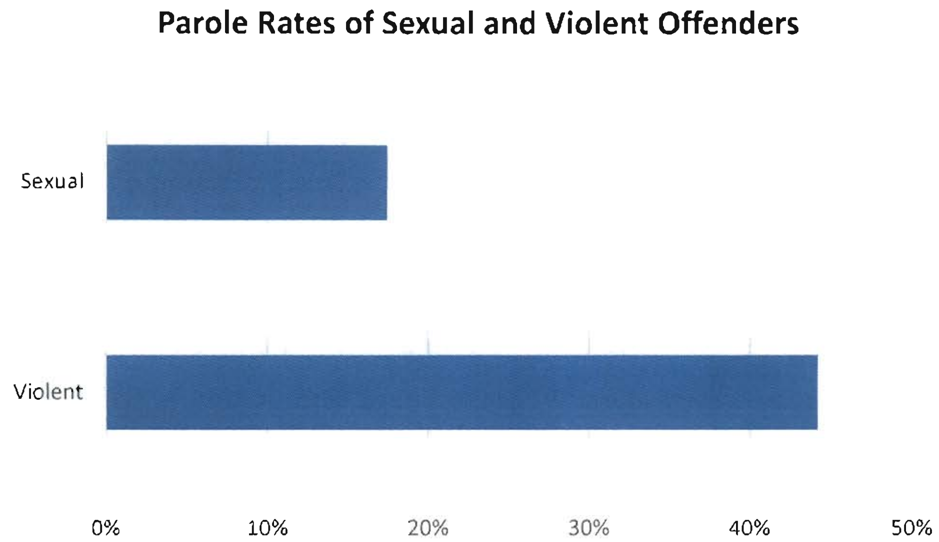
Counts				Percentages		
	Total count	Non-paroled	Paroled	Crime count/total	Non-parole	Parole rate

					rate	
Murder Second Degree	785	381	404	32%	49%	51%
Murder First Degree	701	464	237	28%	66%	34%
Rape First Degree	489	404	85	20%	83%	17%
Sexual Offense 1st Degree	235	196	39	10%	83%	17%
Burglary 1st Degree	63	35	28	3%	56%	44%
Habitual Felon	39	27	12	2%	69%	31%
Armed Robbery	37	10	27	2%	27%	73%
Rape Second Degree	32	20	12	1%	63%	38%
Rape Less than Age 13	16	16	0	1%	100%	0%
AWDWWITKISI	11	4	7	0%	36%	64%
Kidnapping 2nd Degree	9	3	6	0%	33%	67%
Robbery w/Dangerous Weapon	9	3	6	0%	33%	67%
Kidnapping 1st Degree	8	5	3	0%	63%	38%
Sexual Offense 1st Degree w/Child	8	8	0	0%	100%	0%
Sexual Offense 2nd Degree	6	5	1	0%	83%	17%
Burglary 2nd Degree	5	3	2	0%	60%	40%
Arson 2nd Degree	4	1	3	0%	25%	75%
Arson 1st Degree	3	2	1	0%	67%	33%
Trafficking Schedule II	2	1	1	0%	50%	50%
AWDWISI	1	1	0	0%	100%	0%
Felon Unknown	1	0	1	0%	0%	100%
Traffic Opium/Heroin 28 Grams	1	0	1	0%	0%	100%
Trafficking Controlled Substance	1	0	1	0%	0%	100%
TOTAL	**2466**	**1589**	**877**			

When the above crimes are categorized into the types of Violent, Sexual, Property (burglary and arson), Other (habitual felon and felon unknown), and Drug, the differences between parole rates become even more pronounced. It is clear that offenders guilty of serious sexual crimes are paroled at an extremely low rate when compared to violent crime offenders.

	Total	Non-paroled	Paroled	Parole rate
Violent	1561	871	690	44%
Sexual	786	649	137	17%
Property	75	41	34	45%
Drug	4	1	3	75%
Other	40	27	13	33%

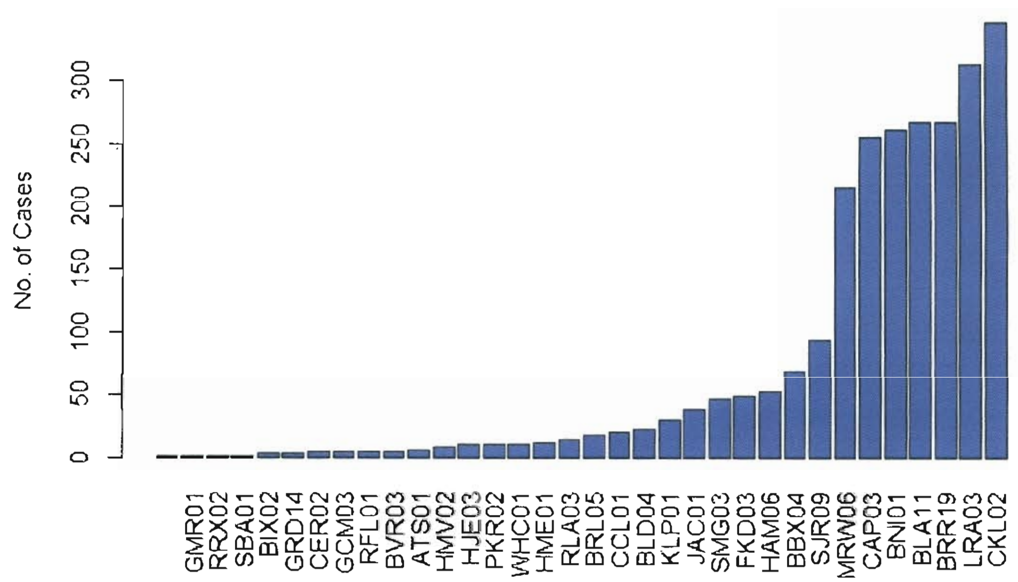
The above data is more clearly visible in the graph below. The graph excludes the "Drug", "Property", and "Other" categories due to the low number of offenders in those categories, which makes them harder to compare. The difference between the parole rates of violent and sexual offenders is hard to miss.



4.1.4 Parole Analyst

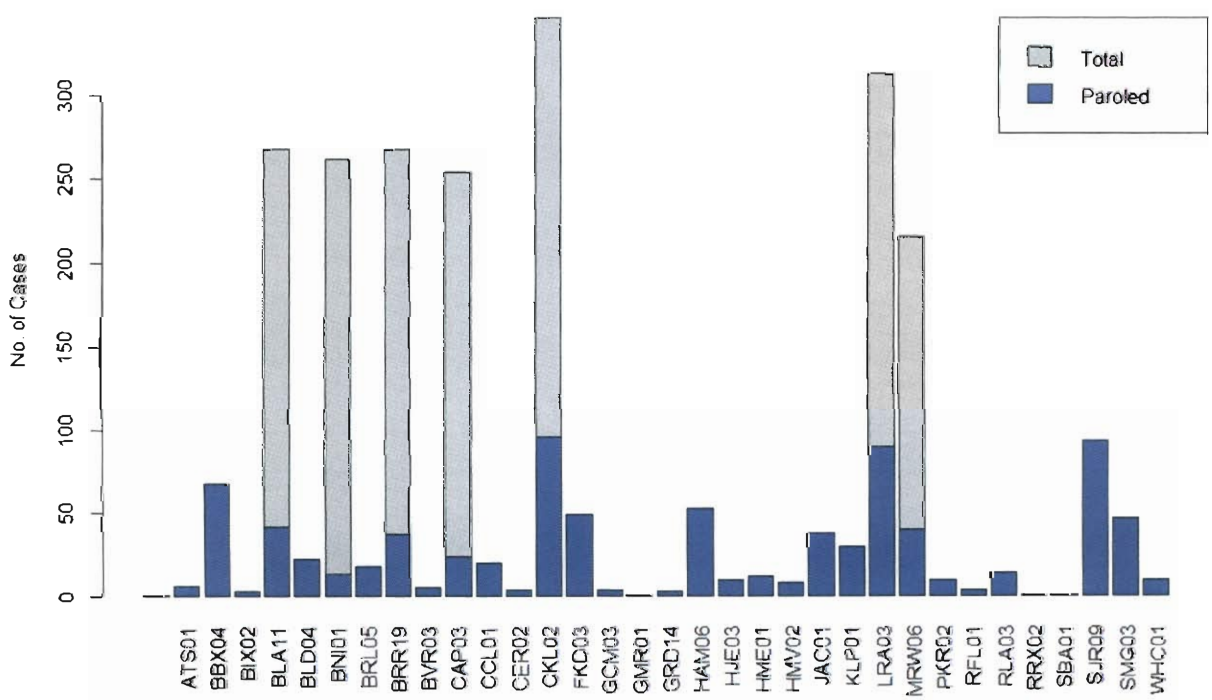
The final element to be examined in this section is the parole analyst field. As mentioned in Section 3.1.4, each parole analyst or parole analyst team is distinguished by an alphanumeric code. The researcher's initial goal was to determine whether the parole analyst as something that varied over time, whether the division of case analysis was divided by case-type or by geography, and whether there was an equal division of cases between analysts. Over time, it became clear that more clarification was needed regarding what each alphanumeric code represented; did it represent an individual analyst, the analyst of a particular region or jurisdiction, or a team of analysts? For the time being, these questions remain unanswered, but a first look does reveal some interesting facts about the division of cases between analysts. It is far from equal, as can be seen in the graph below.

Inmates per Parole Analyst



Moreover, the number and percentage of paroled offenders for each analyst varied widely as well. Some analysts had all of their cases paroled, while others had only a very small portion. For the time being a better understanding of the analyst field is needed before it can be successfully integrated into the model.

Inmates and Parolees per Parole Analyst



5 Model Summary

The previous section looked at the differences between paroled and non-paroled offenders through several different lenses. It is clear that there are some substantial differences between the two groups, in terms of age, race, and sex, disciplinary infractions, and type of crime. Paroled offenders tend to be older and have fewer disciplinary infractions. There are a higher proportion of women being paroled than men, and a higher proportion of white men than black men. Moreover, a much higher proportion of violent offenders are given parole than sex offenders.

Thus there are several possible explanations as to why these differences exist. Perhaps women are less likely to be violent offenders, perhaps whites, on average, are found guilty of fewer disciplinary infractions than blacks. Perhaps rapists also commit more disciplinary infractions. Through the techniques offered above, it is impossible to distill just *to what degree* influential each characteristic is in determining parole decisions. This is the case for regression. Regression allows for the contribution of each element to be isolated and analyzed. Regression allows the researcher to come up with a numerical measure of the overall impact of an inmate's race, for instance, on the outcome. In this case, our dependent variable is "Parole", for which there are two possible values: "Yes" and "No" or 1 and 0. For this sort of variable, standard least squares regression is not a suitable tool, as it can lead to outcomes that outside of the 0:1 range and are therefore uninterpretable. What would a parole value of 3.4 represent, for instance?

5.1 Introduction to Logistic Regression

Using logistic regression allows for results to be interpreted as probabilities, or as odds ratios. Just as in common parlance, statistical probability and odds are a measure of likelihood and comparative likelihood, respectively. Probability can be seen as simply a percent value divided by 100. If something occurs with a probability of 0.98, then there is a 98% chance of it happening. On the other hand, odds attempts to measure how strongly one property is associated with another property. So, for instance, if the researcher is attempting to analyze how an likely an event C is for group A versus group B (type of crime on parole decision, for example), then the odds ratio (OR) between these two is calculated by:

$$OR = \frac{Probability\ of\ C\ for\ group\ A/(1 - Probability\ of\ C\ for\ group\ A)}{Probability\ of\ C\ for\ group\ B\ /(1 - Probability\ of\ C\ for\ group\ B)}$$

If the odds of being granted parole are two times greater for group A than for B, then the odds ratio would be 2, representing that it is twice as likely for group A than for group B. An odds value of 1 would indicate that the results are equally likely. Finally, an odds ratio below 1 would indicate that group A is less likely to experience C than group B.

Logistic regression that probability of getting parole F(x) is related to the explanatory variables x, which are the selected variables, conforms to the following relationship:

$$\mathrm{F(X)} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}}$$

Which equates to a linear relationship with the logarithm off the odds of F(x):

$$ln\frac{F(X)}{1-F(X)} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Thus the coefficients that are being solved for in logistic regression are the β (betas) above, where each beta represents a weight of how important it is in determining the outcome F(x). To interpret the resulting coefficients, they must be exponentiated out of the logarithm, producing a measure of odds.

5.2 A Note on Dealing with Categorical Variables

Categorical variables are those variables that do not have continuous numerical values. For example, race and sex are both categorical variables. Even if we can code a variable for sex/gender as 0 or 1 for male or female, values between 0 and 1 do not have any meaning in this context. Thus categorical variables are treated differently in regression than continuous numerical variables. The categorical variables considered in this research include crime, race, sex, escape attempt flags, and prior flags.

One of the most common ways to deal with different categories, and the way that these variables are dealt with in this report, is through a technique called "dummy variables", which creates new variables to represent the categories. However, each categorical variable must have a "base case" to serve as the standard by which to compare the effect of the categories. For binary variables such as sex, the lack of any variable might represent male and the presence of that variable may represent female. In the context of logistic regression, the coefficient for female would appear in the model, but there would not be any coefficient for male. The interpretation would thus be comparative. The coefficient for the "female" category would represent the expected change in log odds for females as compared to males.

5.3 Results from the Model - Conclusions

In an effort to restrict the research question to demographic categories for which there exists a large enough sample size in the dataset, the researcher performed logistic regression on the portion of the dataset relevant to black and white men only. Thus offenders of other races, and women, were excluded from this analysis to prevent the low number of examples from these groups from interfering with analysis an interpretation.

The first model includes categorical variables for Race, Crime Type, y/n flag for Escapee, y/n flag for Priors, and the numerical variables Infraction Coefficient, Age, and Length of Incarceration. The base case for this model is "White" for Race, "Violent" for Crime Type, and "N" for Escapee and Prior flags.

Model 1	Log Odds	Odds *e^(Log Odds)*	Std. Error	z value	Pr(>|z|)	Significance
(Intercept)	-31.51	**0.00**	339.164	-0.093	0.926	
Race: Black	-0.33	**0.72**	0.134	-2.478	0.013	*
Crime: Drug	14.54	**2071326.83**	706.280	0.021	0.984	
Crime: Other	0.58	**1.78**	0.444	1.304	0.192	
Crime: Property	-0.11	**0.90**	0.336	-0.323	0.747	
Crime: Sex	-1.47	**0.23**	0.153	-9.580	0.000	***

Escapee	-1.42	**0.24**	0.263	-5.393	0.000	***
Priors	28.86	**3401931694859.50**	339.163	0.085	0.932	
Infraction Coefficient	-0.10	**0.90**	0.011	-8.935	0.000	***
Age	0.37	**1.45**	0.027	13.912	0.000	***
Length of Incarceration	0.00	**1.00**	0.000	6.928	0.000	***

This model contains two fields with absurdly large odds, both in field which have a very high standard error and low significance. Thus excluding these fields as well is a reasonable step. The second model includes only the Crime Types of "Violent", "Sex", "Property", and "Other", and does not include the Priors flag.

Model 2	**Log Odds**	**Odds *exp(Log Odds)***	**Std. Error**	**z value**	**Pr(>|z|)**	**Significance**
(Intercept)	-3.75	**0.02**	0.3570	-10.4910	0.0000	
Race: Black	-0.17	**0.85**	0.1117	-1.5060	0.1321	
Crime: Other	1.03	**2.81**	0.4254	2.4248	0.0153	*
Crime: Property	0.12	**1.13**	0.2974	0.3962	0.6920	
Crime: Sex	-1.55	**0.21**	0.1300	-11.9547	0.0000	***
Escapee	-1.53	**0.22**	0.2453	-6.2508	0.0000	***
Infraction Coefficient	-0.09	**0.91**	0.0109	-8.3591	0.0000	***
Age	0.26	**1.30**	0.0172	15.3718	0.0000	***
Length of Incarceration	0.00	**1.00**	0.0000	8.2225	0.0000	***

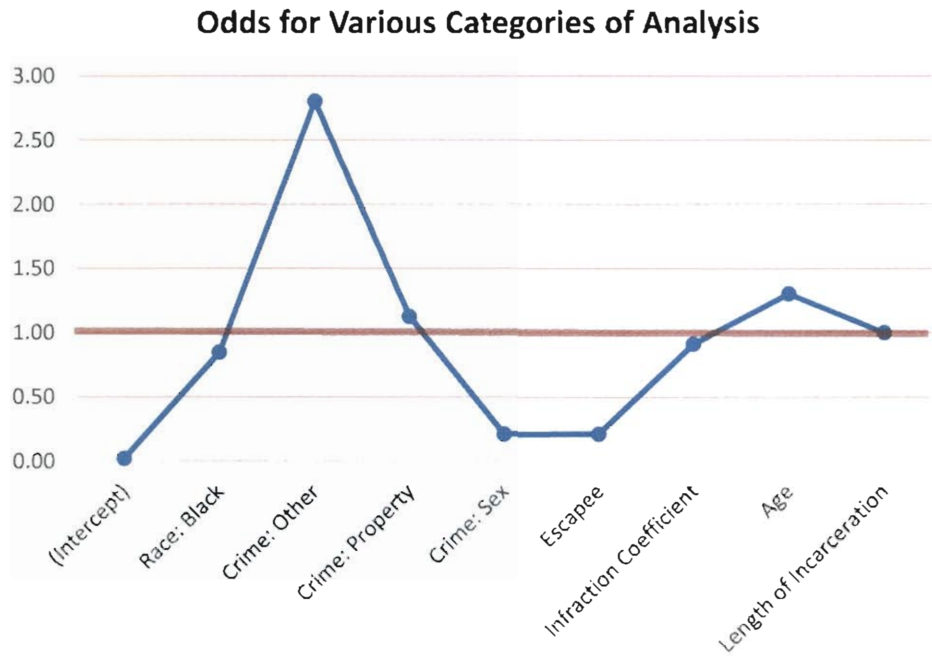
Using this model, we see that the significance rating for most of our coefficients is quite high. In addition, with a higher standard error, the Race variable has lost some of its significance, although it still seems to be within reasonable enough bounds to be included in the model. The same also applied for the Property crime category.

Interpreting these results is a somewhat meticulous process at it depends on the variable in question. For the categorical variables, the base cases must be kept in mind. For race, according to this model, black inmates have 0.85 odds of being paroled compared to whites. This would mean that, based on race alone, blacks have a lower likelihood of being paroled than whites, all else being equal. Considering the complexity of this analysis, and the low significance of Race as a variable in this model, there is certainly not conclusive evidence of systematic racism. However, the possibility certainly cannot be ruled out based on this analysis.

The crime variable is both significant and large for many categories. It is clear that, compared against the base case of violent crime, sex offenders are *significantly* less likely to be paroled. On the other hand, perpetrators of property crimes (which include burglary and arson in this model) are only slightly more

likely to be paroled than violent offenders. Finally, offenders in the other category appear to be granted parole with a significantly higher probability than violent offenders.

The last categorical variable included in this model, the escape flag, also seems to be one of great importance. Both significant statistically and in magnitude, it appears that those that attempt escape are significantly less likely to be granted parole.



For the numerical variables, a slightly different interpretation must be kept in mind. While the categorical variables are an either/or scenario, the numerical variables have a large range. Thus while the odds for these variables are lower than for many of the categorical variables, these apply to the log odds *per unit* of the numerical variable. Therefore, while infraction coefficient may only have an odds of 0.91, the logarithm of that will be multiplied by a variable which ranges from 0 to over 800. Thus a while 0.91 may seem relatively close to 1 (which would indicate that infraction coefficient is of no predictive significance), the range of the variable makes this component matter. Prisoners with high infraction coefficients are less likely to be paroled.

Age is another variable that has an aggregate affect depending on the magnitude. This model shows that older offenders are more likely to be paroled, and that the effect of age is quite significant.

Finally, somewhat contrary to the researcher's expectation, the length of an offender's incarceration seems to have no impact on whether or not they will be paroled. Merely being in prison longer is not enough to increase parole likelihood.

6 Appendix

Infraction Severity Scale

Offense	Severity	Offense	Severity
ACTIVE RIOTER	2	LEGAL ASSISTANCE	1
ASSAULT ANOTHER W-SEX INT	3	LOCK TAMPERING	2
ASSAULT INMATE W-SEX INT	3	MISUSE MEDICINE	1
ASSAULT PERSON W-WEAPON	2	MISUSE SUPPLIES	1
ASSAULT STAFF W-SEX INT	3	MISUSE-UNAUTH-USE PHONE-MAIL	1
ASSAULT STAFF W-WEAPON	3	NEGLIGENTLY PERFORM DUTIES	1
ASSAULT STAFF-INSTIGATE-PROVOK	3	NO THREAT CONTRABAND	1
ASSAULT STAFF-THROWING LIQUIDS	2	OFFER-ACCEPT BRIBE ANOTHER	1
ASSLT INMATE-THROWING LIQUIDS	2	OFFER-ACCEPT BRIBE STAFF	1
ASSLT OTHER W-UNLIKELY INJ	2	POSS AUDIO-VIDEO-IMAGE DEVICE	1
ASSLT STAFF W-UNLIKELY INJ	3	POSS MONEY-UNAUTHORIZED FUNDS	1
ATTEMPT CLASS A OFFENSE	1	POSSESS EXCESS STAMPS	1
ATTEMPT CLASS B OFFENSE	2	POSSESSION MONEY	1
ATTEMPT CLASS C OFFENSE	2	PRE-OPUS CONVERSION	1
ATTEMPT CLASS D OFFENSE	1	PROFANE LANGUAGE	1
ATTEMPT CLASS E OFFENSE	1	PROPERTY TAMPERING	1
BARTER-TRADE-LOAN MONEY	1	PROVOKE ASSAULT	1
BED VIOLATION	1	REFUSE SUBMIT-DRUG-BREATH TEST	1
CREATE OFFENSIVE CONDITION	1	INVOLVEMENT W-GANG OR STG	2
DAMAGE STATE-ANOTHERS PROPERTY	1	SELF INJURY	1
DETONATING EXPLOSIVES	3	SELL-MISUSE MEDICATION	1
DISOBEY ORDER	1	SET A FIRE	3
ESCAPE	3	SEXUAL ACT	2
EXTORTION-STRONG ARM	2	SUBSTANCE POSSESSION	2
FAKE ILLNESS	1	TAKING HOSTAGE(S)	3
FALSE ALLEGATIONS ON STAFF	1	THEFT CANTEEN INV-CASH	2
FALSE INFO CLASS A OFFENSE	2	THEFT OF PROPERTY	1
FALSE INFO CLASS B OFFENSE	2	THREATEN TO HARM-INJURE STAFF	2
FIGHT W-WEAPON OR REQ.OUT.MED	2	UNAUTHORIZED FUNDS	1
FIGHTING	2	UNAUTHORIZED LEAVE	2
FLOOD CELL	1	UNAUTHORIZED LOCATION	1
FORGERY	2	UNAUTHORIZED TOBACCO USE	1
GAMBLING	1	UNCLEAN BODY	1
HIGH RISK ACT	2	UNKEMPT ROOM	1
ILLEGAL CLOTH-LINEN-SHEETS	1	UNWANTED COMMUNICATE W-VICTIMS	2

INHALE SUBSTANCE	1	VERBAL THREAT	1
INTERFERE W-STAFF	1	VIOLATE NC LAW	1
INVOLVEMENT W-GANG OR STG	2	WEAPON POSSESSION	2
LEAVEQUIT COMM BASED PROGRAM	1	WRK STOPPAGE-COMM. WORK CREW	2

Author:	**Date:**
Bryan Gilbert Davis	August 20, 2014
Bryan Davis	

Bryan Davis
Email: bryan.g.davis@gmail.com Cell: 828-318-4279
Currently based in Chapel Hill, North Carolina

Education	**University of North Carolina Chapel Hill** Masters of Science in Statistics and Operations Research Expected graduation: May 2015 **University of North Carolina Chapel Hill** Bachelors of Arts in History and Asian Studies Graduated with distinction in May 2008
Skills	• Programming Java, Python, and R • Statistical analysis and machine learning • Data structure and algorithm design and analysis • Experience with SQL query building and database design • Fluent in Mandarin Chinese with translation and interpreting experience

---

Work experience:

Information Technology	**Graduate Student Technical Support at UNC Research Computing** (its.unc.edu/research) • Answer technical queries on operation of research software on UNC's Linux supercomputer clusters • Develop technical support documentation to guide users on how to use the clusters	Chapel Hill, North Carolina January 2014 to present
Economic research and analysis	**Wind Industry Analyst at Bloomberg New Energy Finance (BNEF)** (www.bnef.com) • Created supply and demand models China's wind power industry. • Developed data-driven research reports for BNEF's more-than 500 clients. • Personally assisted dozens of individual clients in managing and understanding their multi-million dollar relationships with Chinese suppliers and customers.	Beijing, China August 2011 to June 2012
	Wind Industry Analyst at Azure International (www.azure-international.com) • Oversaw hundreds of pages of research on China's wind power industry • Developed models of supply networks and production capacity • Wrote database queries to address specific research questions	Beijing, China September 2010 to July 2011

Honors and Awards	**Best Visualization** at Duke Datafest 2014 **Foreign Language Area Studies Scholarship** 2014 **Frances L. Phillips Travel Scholarship** 2008